

So, you need simulated data?

MC and SAM

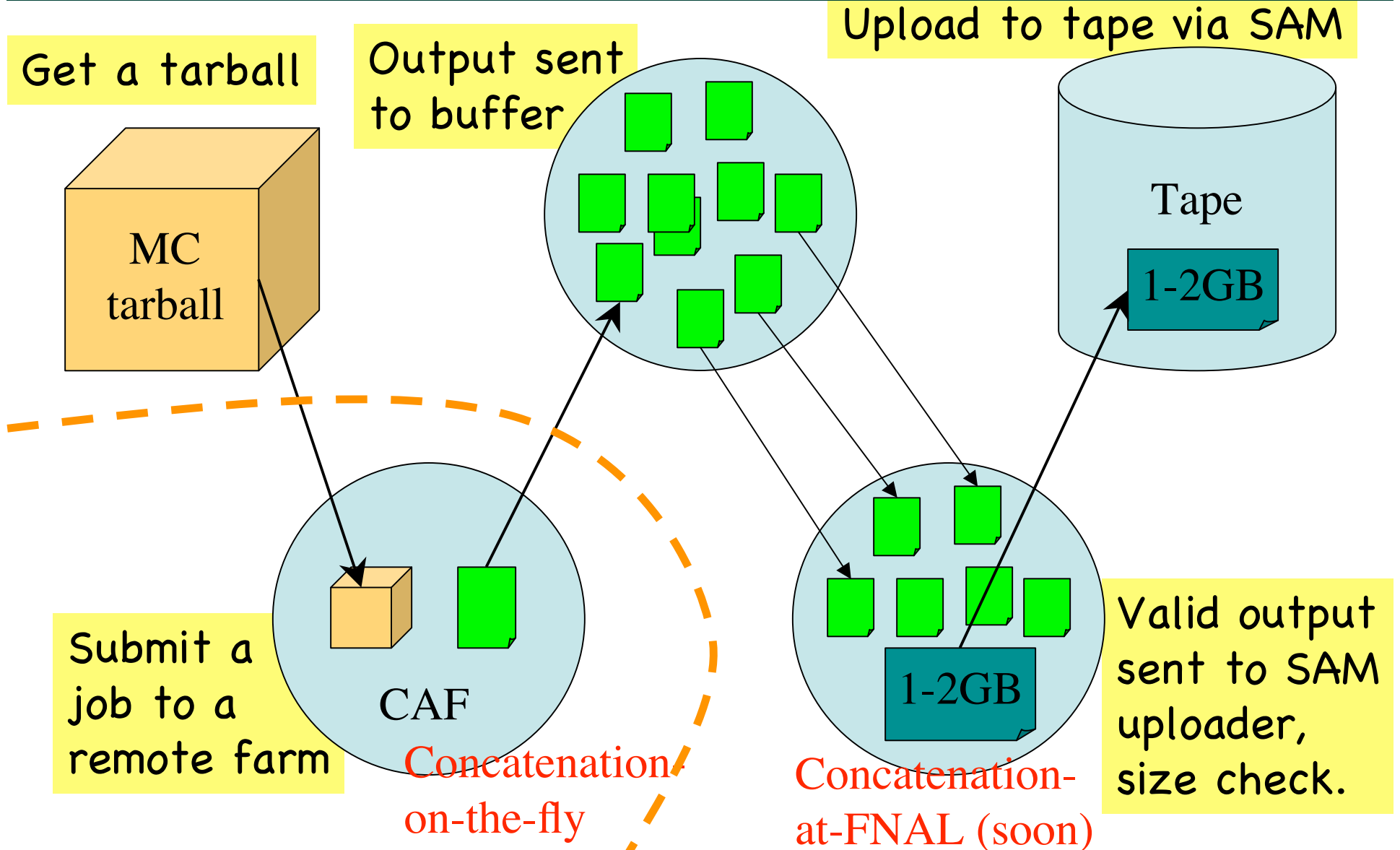
Rob Harr
for MC
production



Outline

- Overview of MC generation process
- How your MC data gets from the farm to tape.
- Recent statistics on MC upload with SAM
- Description of and planned improvements to MC production
- Acknowledgements:
 - DH group: Benjamin, Borgatti, Genser, Zhang
 - MC group: Beauchemin, S.Y. Jun, Warburton
 - MC reps: Hsu, Ivanov, Whiteson, ...

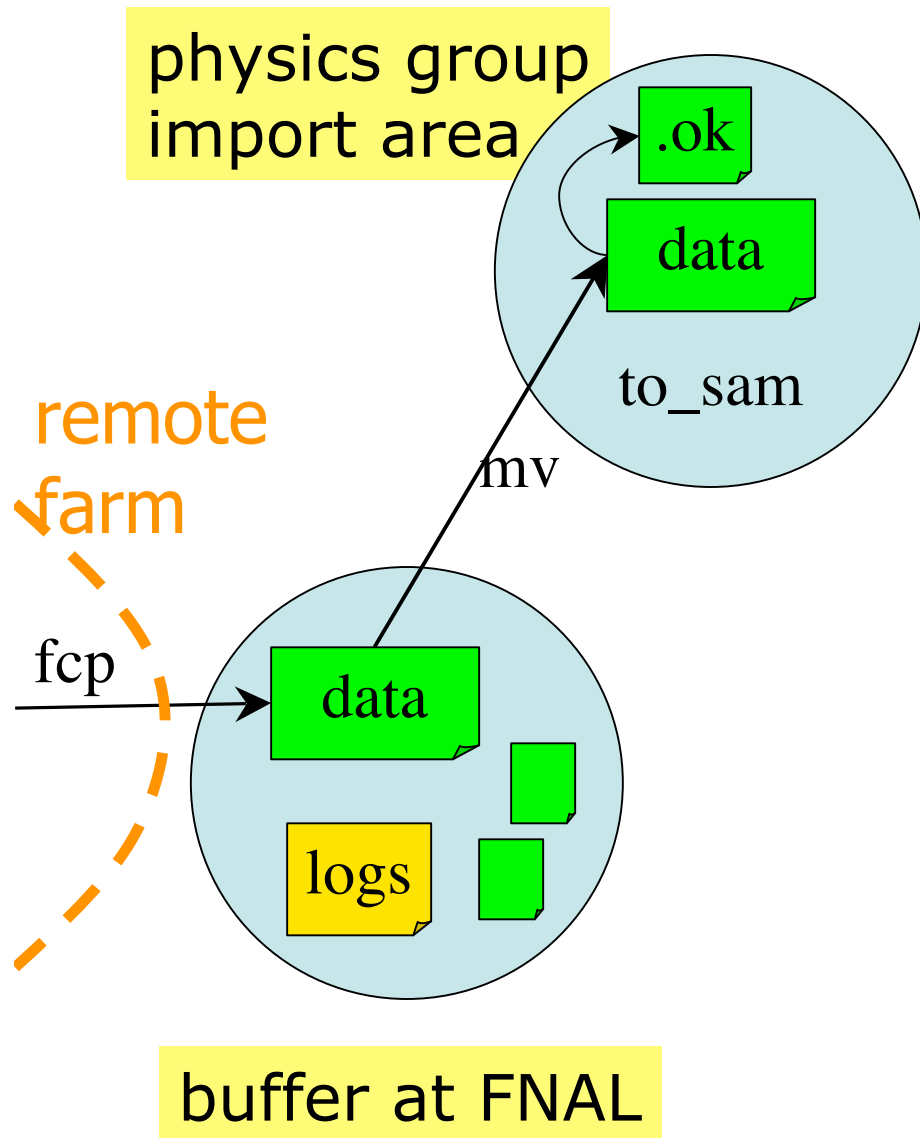
Overview of MC Generation



Getting Your Files to Tape

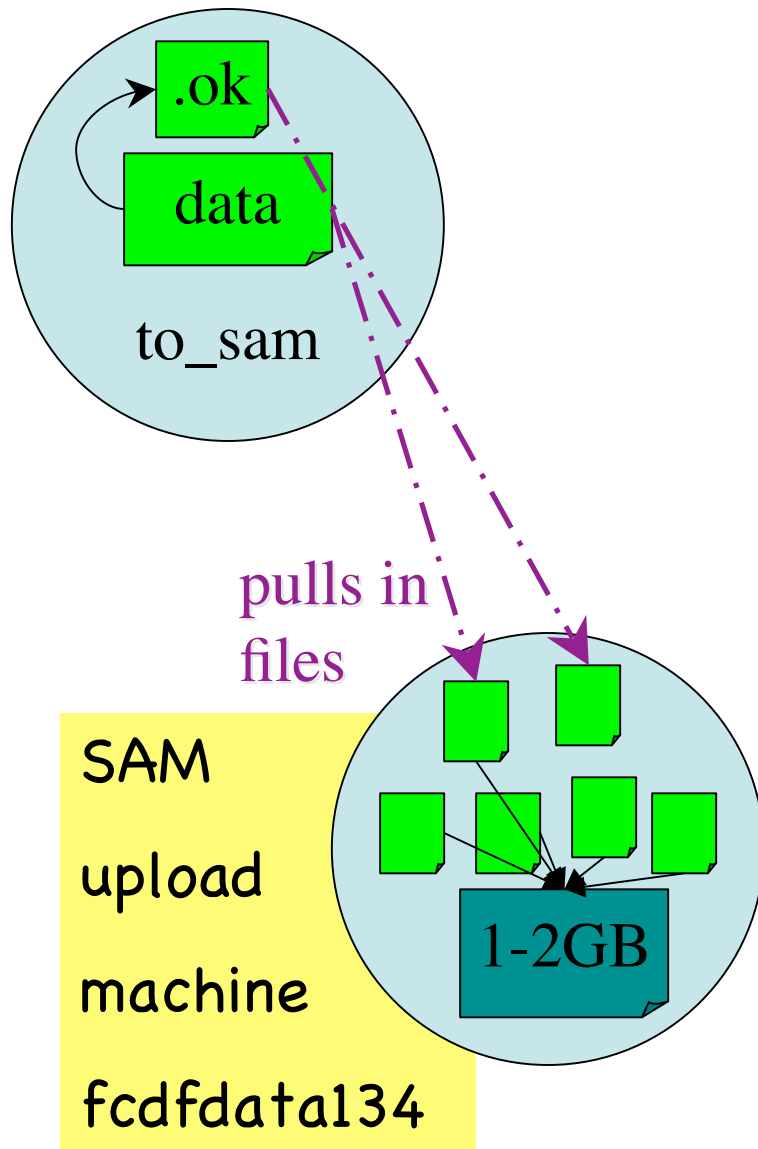
- Declare your dataset to SAM
 - MC reps. have privilege to do this
- Submit your job to a remote farm
- Direct job output to a buffer area at FNAL
 - buffers controlled by physics groups
 - consult MC rep for access to physics group buffer
 - “official” production script uses fcp to return job output.
- Check output files and logs.

From Farm to Physics Buffer



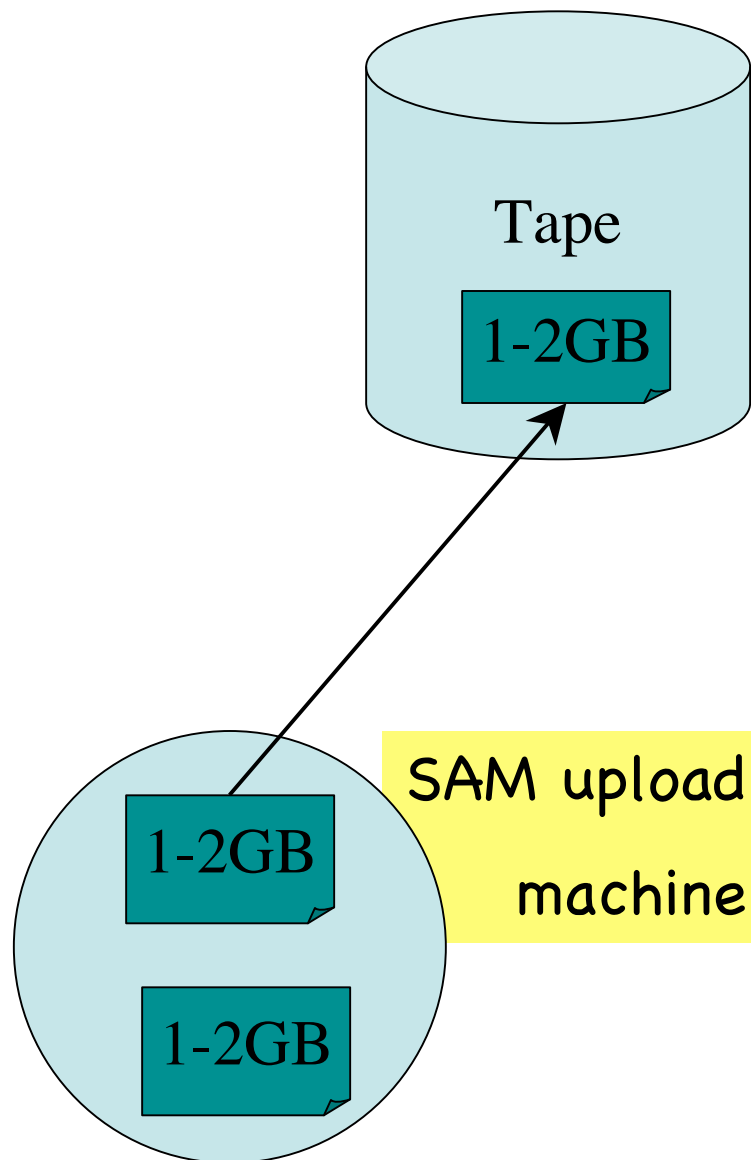
- Farm output (data and log files) fcp'd to specified buffer area
- User checks MC data and log files for validity
 - md5 check sum
 - log file check script (new)
- If ready for SAM upload, move data files to "import area" (if needed) along with .ok file
 - import area is a designated directory, usually "to_sam"
- Archive log files

From Buffer to SAM Machine



- The .ok files signal that data is ready to upload to tape.
 - Default contains email
 - Extra info in future
- A cron job pulls data and .ok files to the SAM upload machine
 - fcdldata134 (133 is backup)
- Delete files from buffer after successful import.
- monitoring: http://www-cdf.fnal.gov/~danielw/data_to_tape/report_fcdldata134_N.html (N=1,2,3)

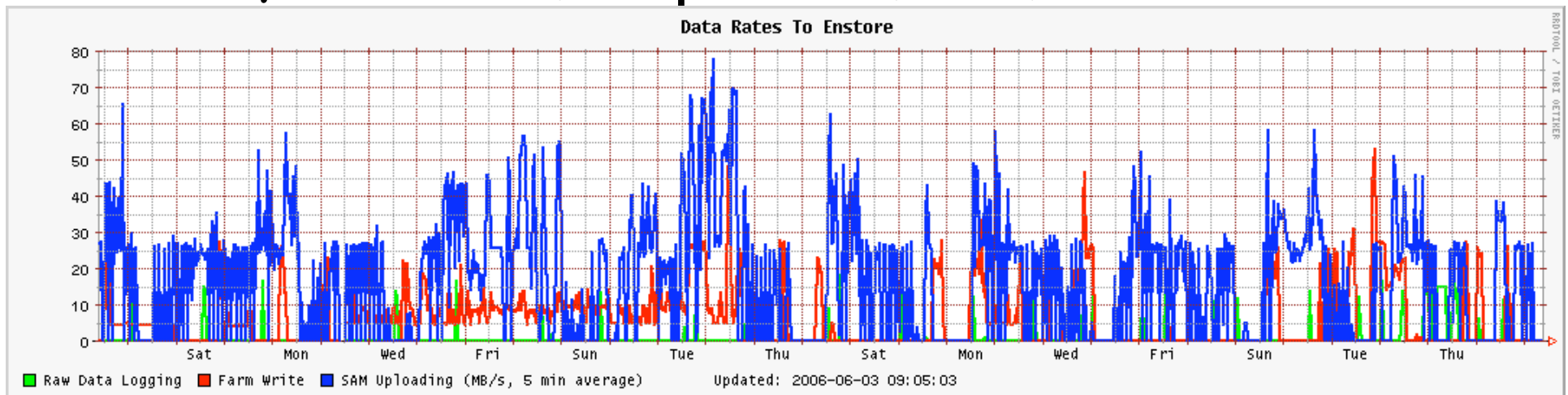
SAM Machine to Tape



- SAM station handles the upload to tape.
- Want file sizes between 1 and 2 GB for efficient tape usage.
 - Temporary 200MB limit
 - New: Implement concatenation at FNAL before upload
- Error checking
 - size, dataset definition, name, SAM running, glitches, ...
- Delete files after successful upload

SAM MC Upload Statistics

- Started using SAM for MC in January, switched all MC upload to SAM in April, disabled DFC uploads on May 31.
- Files uploaded in May: 16454
- Files uploaded since Jan.: 48742
- 25MB/s rate to tape is common



What's in the MC Tarball

- Everything including the kitchen sink
 - all executables: cdfSim, TrigSim++, Production, EdmEventLister, mysql, and sqlplus
 - Frontier linked by default
 - all shared libraries needed by these exe's
 - including system libraries, with a few exceptions
 - standard scripts to run jobs
 - lots of standard tcl's
 - data files: B field map, decay tables, ...
- Usable on grid sites lacking cdfsoft!

Where to Find a Tarball

- “Official” tarballs provided by the MC production group, linked to “MC How-To” page
 - <http://www-cdf.fnal.gov/tiki/tiki-index.php?page=MC%20Production%20How%20To%20Page>
 - standard, stripped, and maxopt versions
- Some power users have created customized versions with extras: topntuple, alpgen, ...
 - ask group MC rep. or MC production group
- Build instructions on “MC How-To” page
 - please check with MC production group because recent fixes require some special procedures.
 - users are encouraged to use “official” tarballs

How to Submit Your Job

- Create a dataset definition file
 - instructions on MC How-To page
- Extract mcProduction/scripts/submit_MCPProd
- Choose a farm
 - new scheme of CAF clusters: NAmCAF is first
 - will choose best farm for job
- Submit job following instructions on How-To page for submit_MCPProd.
- Please report bugs to cdf_mc_ops@fnal.gov
 - save log files and other output

Improvements in MC Production

- Efforts to improve reliability, speed, and ease of use.
- Higher priority MCProd group for MC reps.
- Maintain usability on computing resources.
 - all-in-one tarball
 - average user can use pre-built “official” tarball
- Frontier is the default for DB access (6.1.4)
- Formalize procedure for checking a new tarball and documentation for users.
- Maxopt tarball: ongoing validation.
 - 30% to 50% faster according to users

More Improvements

- Concatenation-at-FNAL: shorter job times, easier to recover failed sections.
- Log file check script – R. Snider & C. Plager
- md5 checksum on output files – S. Hsu
- split tarball submission scheme – J. Galyardt & I. Sfiligoi
 - faster submission, less strain on headnodes
- merging B MC with high-pT MC – M. Campanelli
 - will probably yield new scheme for modifying decay tables

Summary

- Switched MC uploads from DFC to SAM; only minor problems encountered
- Multi-step scheme to move data from farm to tape
- MC tarball designed to be widely usable
- Implementing improvements for robustness and ease of use
- Many people have contributed to implementing and testing MC production